See related commentary on pg 11

Can We Rely on the Dermatology Life Quality Index as a Measure of the Impact of Psoriasis or Atopic Dermatitis?

James Twiss[1], David M. Meads[1], Elizabeth P. Preston[1], Sigrid R. Crawford[1] and Stephen P. McKenna[1]

The Dermatology Life Quality Index (DLQI) is a widely used health-related quality of life measure. However, little research has been conducted on its dimensionality. The objectives of the current study were to apply Rasch analysis to DLQI data to determine whether the scale is unidimensional, to assess its measurement properties, test the response format, and determine whether the measure exhibits differential item functioning (DIF) by disease (atopic dermatitis versus psoriasis), gender, or age group. The results show that there were several problems with the scale, including misfitting items, DIF by disease, age, and gender, disordered response thresholds, and inadequate measurement of patients with mild illness. As the DLQI did not benefit from the application of Rasch analysis in its development, it is argued that a new measure of disability related to dermatological disease is required. Such a measure should use a coherent measurement model and ensure that items are relevant to all potential respondents. The current use of the DLQI as a guide to treatment selection is of concern, given its inadequate measurement properties.

Journal of Investigative Dermatology (2012) **132,** 76–84; doi:10.1038/jid.2011.238; published online 1 September 2011

INTRODUCTION

The Dermatology Life Quality Index (DLQI; Finlay and Khan, 1994) is the most commonly used patient-reported outcome measure in dermatology. It was designed to be a generic dermatology health-related quality of life (HRQL) questionnaire and has been translated into 55 languages. The measure has been used in many clinical studies in over 30 skin conditions, but primarily in studies of psoriasis and atopic dermatitis (AD; Lewis and Finlay, 2004; Basra *et al.*, 2008). In the United Kingdom, the DLQI is used to determine whether patients are eligible to receive biological interventions for psoriasis (Smith *et al.*, 2005, 2009; NICE, 2008a, b, 2009).

Although the classical psychometric properties (including test retest reliability, internal consistency, and construct validity) of the questionnaire have generally been found to be adequate, more detailed analyses using Rasch analysis have highlighted several problems with the scale (Nijsten *et al.*, 2006a, 2007). Rasch analysis (Rasch, 1960) is now seen as the method of choice for the development and improvement of questionnaires as it has several advantages over Classical Test Theory approaches such as factor analysis (Wright, 1996; Wright and Tennant, 1996; Luquet *et al.*, 2001; Prieto *et al.*, 2003; Tennant *et al.*, 2004; Waugh and Chapman, 2005; Nijsten *et al.*, 2006a).

DLQI scores for individuals with different types of skin conditions are often combined in research studies (Papoutsaki *et al.*, 2007; Schmitt *et al.*, 2007; Potocka *et al.*, 2008, 2009; Quandt *et al.*, 2008; Ludwig *et al.*, 2009). Despite this, no previous research has assessed whether it is justifiable to combine DLQI data in this way. One way in which this can be determined is by the use of differential item functioning (DIF) analysis within the Rasch framework (Holland and Wainer, 1993; Brodersen *et al.*, 2010).

The aim of the present study was to reinvestigate the scaling properties of the DLQI in combined samples of psoriasis and AD patients. The study also investigated the scaling properties of the DLQI in psoriasis and AD populations separately. The Rasch model was used to determine:

- whether the scale was unidimensional,
- the level of trait (HRQL) covered by the scale,
- how well items were targeted to the populations,
- whether the response categories were working in a logical way,
- whether the DLQI was free from DIF by age and gender, and
- whether it is justifiable to compare scores of psoriasis and AD samples using the DLQI.

RESULTS

Demographic details of the samples are shown in Table 1. The two samples were well matched in terms of age and

[1]*Galen Research, Manchester, UK*

Correspondence: James Twiss, Galen Research, Enterprise House, Manchester Science Park, Lloyd Street North, Manchester M15 6SE, UK.
E-mail: jtwiss@galen-research.com

Abbreviations: AD, atopic dermatitis; DIF, differential item functioning; DLQI, Dermatology Life Quality Index; HRQL, health-related quality of life; PSI, Person Separation Index; QoL, quality of life

Received 3 September 2010; revised 5 May 2011; accepted 5 June 2011; published online 1 September 2011

Table 1. Sample characteristics

	Psoriasis (*n*=146)	Atopic dermatitis (*n*=146)
Gender (%)		
Male	73 (50.0)	73 (50.0)
Female	73 (50.0)	73 (50.0)
Age (years)		
Mean (SD)	44.4 (14.7)	45.5 (16.6)
Range	66.0 (17.0–83.0)	62.0 (20.0–82.0)
Duration (years)		
Mean (SD)	20.9 (13.5)	28.2 (17.5)
Range	67.0 (2.0–69.0)	76.0 (0.0–76.0)
Perceived general health (%)		
Excellent	16 (11.0)	15 (10.3)
Good	68 (46.6)	96 (65.8)
Fair	44 (30.1)	32 (21.9)
Poor	17 (11.6)	3 (2.1)
Perceived illness severity (%)		
Mild	38 (26.0)	60 (41.1)
Moderate	57 (39.0)	57 (39.0)
Quite severe	43 (29.5)	26 (17.8)
Very severe	6 (4.1)	3 (2.1)
DLQI score		
Mean (SD)	29.2 (22.3)	20.7 (16.7)
Median (IQR)	23.3 (10.0–40.8)	15.5 (10.0–26.7)
Range	96.7 (0.0–96.7)	86.7 (0.0–86.7)
Marital status (%)		
Married/living as	89 (61.0)	95 (65.1)
Single	57 (39.0)	50 (34.2)

Abbreviations: DLQI, Dermatology Life Quality Index; IQR, interquartile range.

gender. AD patients had had their condition for longer ($P<0.001$). DLQI scores were higher for the psoriasis group ($P<0.001$), which also reported worse perceived general health ($P=0.001$) and perceived illness severity ($P=0.018$).

Item frequencies

DLQI raw score item responses for the combined samples are shown in Figure 1. The figure indicates how frequently each item response was used. A high proportion of participants responded ''not relevant'' to items 6, 7, and 9.

Rasch analysis

The initial likelihood ratio test, used to determine the most appropriate Rasch model, was statistically significant, supporting the use of the partial credit model for the analyses ($P<0.001$).

Rasch analysis of the combined sample—initial run

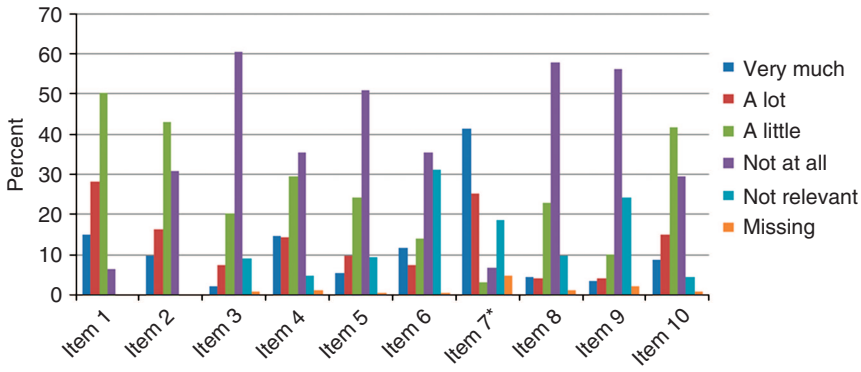
Initial overall scale fit. Table 2 shows the Rasch overall fit statistics for the combined sample. The DLQI misfit the Rasch model (item–trait interaction, $P=0.010$). This confirms that the DLQI does not form a unidimensional measure of health-related quality of life (HRQL) for combined psoriasis and AD samples. In addition, the mean item residual statistic was low and further indicative of misfit to the Rasch model. This suggests that there are problems with individual items in the scale. The Person Separation Index (PSI) indicated that the DLQI had adequate internal reliability. One pair of items (items 1 and 6; see item descriptions listed in Table 3) had a high residual correlation (0.312), suggesting a response dependency between the items. In this case, answers to one item had too strong an association to answers to the other.

Initial item fit. Table 3 reports the fit statistics for the items in the DLQI. Items 2 ($P=0.005$), 5 ($P=0.013$), and 7 ($P=0.013$) showed misfit to the model. Item 5 also had a low-fit residual (<-2.5), suggesting that it was redundant.

Differential item functioning. Table 3 also shows the items with significant DIF. Figure 2 illustrates the DIF by disease found for item 7. It plots the observed item characteristic curve for both disease groups on the same logit scale and against the predicted ''S''-shaped curve. At every level of severity from left to right (mild to severe), AD patients were more likely to affirm the item, indicating that the item works differently in each disease.

Item response thresholds. Response thresholds were disordered for items 4, 6, 7, 8, and 9, showing that the response formats for these items did not work logically. Further investigation of these showed that response options 2 (''a lot'') and 3 (''very much'') were disordered for items 4, 7, and 8. Response options 1 (''a little''), 2 (''a lot''), and 3 (''very much'') were disordered for items 6 and 9, indicating that a dichotomous response format would be more appropriate for these two items.

Item locations and logit coverage. The DLQI item map for the combined sample is shown in Figure 3. The location of each person and each item response on the same logit scale is illustrated. The map shows that items are bunched around the middle of the logit scale, and that there are few items covering the mild end. This indicates that the scale does not work well with individuals who have mild disease. For example, the scale would not be able to show improvements in scores for patients with milder disease who benefit from treatment. The map also shows some redundancy in the scale.



*Item 7 is a two-part question, it first asks whether work or study has been prevented (yes/no) and then (if 'No') to what degree the skin condition has been a problem at work/study ('A lot', 'A little', or 'Not at all'). If participants answer 'Yes' to the first part of the question they are scored the same as 'Very much'.

Figure 1. Dermatology Life Quality Index (DLQI) raw-score item frequencies. This shows the percentage of patients who responded with each response option for each question of the DLQI. Information is based on the raw scores.

Table 2. Rasch analysis overall fit statistics

Sample	Analysis number	Description	*n*	Item–person interaction	PSI	Items		Persons	
						Mean	SD	Mean	SD
Combined	Analysis 1	All of sample	282	0.010	0.85	−0.81	0.98	−0.30	0.81
Combined	Analysis 2	Delete two misfitting participants	280	0.011	0.79	−0.86	0.97	−0.30	0.76
Combined	Analysis 3	Collapse responses 2 and 3 for items 4, 6, 7, 8, and 9	280	0.040	0.72	−0.79	0.88	−0.28	0.78
Combined	Analysis 4	Collapse responses 2, 3, and 4 for items 6 and 9	279	0.063	0.78	−0.72	0.80	−0.27	0.83
Combined	Analysis 5	Delete item 2	279	0.021	0.72	−0.89	0.85	−0.30	0.81
Combined	Analysis 6	Reinstate item 2 and split items 4 and 7 by disease	279	0.076	0.78	−0.60	0.75	−0.27	0.82
Combined	Analysis 7	Split item 10 by age	279	0.185	0.78	−0.56	0.75	−0.27	0.83
Psoriasis	Analysis 1	All of sample	141	0.005	0.82	−0.23	0.92	−0.24	0.85
Atopic dermatitis	Analysis 1	All of sample	141	0.460	0.63	−0.64	0.66	−0.30	0.78

Abbreviation: PSI, Person Separation Index.

Rasch analysis of the combined sample—refinement of the scale
Several changes were made to the scale in an attempt to improve its measurement properties. The effects of the changes to the scale are summarized in Table 2.

In Analysis 2, participants showing misfit to the model were removed. This made little change to the overall model fit and the scale continued to show the problems identified in the first analysis. In Analysis 3, responses of ''a lot'' and ''very much'' were combined for items 4, 6, 7, 8, and 9 to correct for the problems with these responses. This resulted in a marginal improvement in the overall scale fit statistics and the response problems for items 4, 7, and 8 were resolved. However, response problems remained for items 6 and 9. It appeared that a dichotomous response format would be more appropriate for these two items. Consequently, for Analysis 4, the responses ''a little'', ''a lot'', and ''very much'' were combined for items 6 and 9. Overall fit statistics again improved marginally. The measure continued to show problems related to item misfit and DIF by age, gender, and disease. In particular, item 2 continued to show significant misfit ($P = 0.007$). This item was deleted for Analysis 5, which actually caused the overall fit statistics for the scale to deteriorate. Therefore, item 2 was reinstated.

For Analyses 6 and 7, the items showing the greatest level of DIF were split. In Analysis 6, items 4 and 7 were split in order to become separate items for psoriasis and AD patients (as they exhibited DIF associated with disease). In Analysis 7, item 10 (where there was DIF by age) was also split. Following Analysis 7, the overall fit statistics for the DLQI were improved. Despite this, the item fit residuals suggested that there were continuing problems with the items. Item 2 continued to show misfit to the model ($P = 0.009$), and several items continued to show low-level DIF by age, gender, and disease. Furthermore, these changes were unable to overcome the problems associated with the scale's poor coverage of mild impairment.

Table 3. Individual item fit statistics and DIF

Item description	Location	Fit residual	χ^2	*P*-value	Uniform DIF	Non-uniform DIF
(1) Itchy, sore, painful, or stinging	−1.61	−0.03	3.38	0.496	Disease (*P*=0.04)	
(2) Embarrassment/self-consciousness	−0.52	−1.97	15.00	0.005	Age group (*P*=0.003)/ gender (*P*=0.010)	
(3) Interferes with shopping/looking after home/garden	0.99	−1.18	3.75	0.441	Disease (*P*=0.008)	Disease (*P*=0.03)
(4) Influences choice of clothes	−0.54	−0.14	4.43	0.350	Gender (*P*=0.002)/ disease (*P*=0.014)	
(5) Affects social/leisure activities	0.32	−2.69	12.73	0.013	Disease (*P*=0.025)	
(6) Affects ability to do sports	−0.02	0.29	3.00	0.558	Gender (*P*=0.006)	Gender (*P*=0.009)
(7) Prevents working/studying	0.25	0.23	12.59	0.013	Gender (*P*=0.007)/ disease (*P*<0.001)	
(8) Creates problems with partner/close friends/relatives	0.59	−0.89	2.12	0.715		
(9) Causes sexual difficulties	0.89	−1.23	2.20	0.698		
(10) Problem with treatment	−0.36	−0.47	3.48	0.481	Age group (*P*<0.001)	

Abbreviation: DIF, differential item functioning.

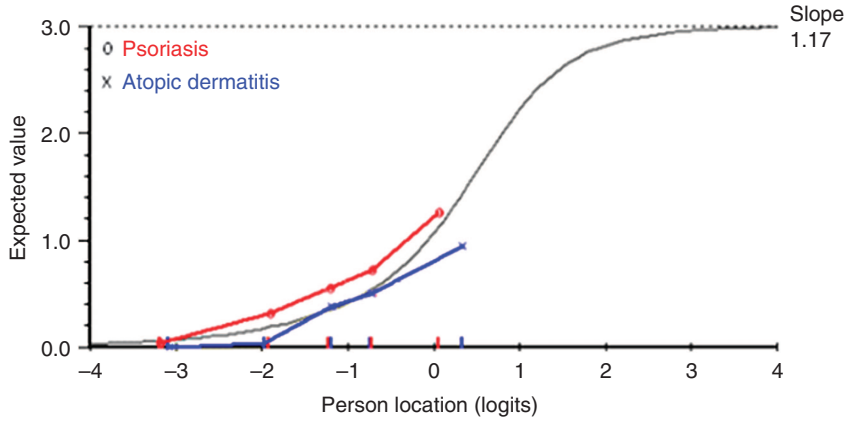


Figure 2. Differential item functioning by disease for work/study item. This shows the item characteristic curve (ICC) for item 7 of the Dermatology Life Quality Index. In Item Response Theory, an ICC describes the relationship between a latent ability and the performance on a test item. The curved line represents the expected scores for this item. The actual scores for the item are shown for psoriasis and atopic dermatitis patients separately.

Rasch analysis of the psoriasis and AD samples separately

Further analyses were conducted to assess whether the DLQI fit the Rasch model for the psoriasis and AD samples separately. Overall fit statistics for the two samples are shown in Table 2.

Psoriasis

Fit statistics showed that the DLQI scores for the psoriasis sample misfit the Rasch model (item–trait interaction = 0.005). Thus, the DLQI was not a unidimensional measure of HRQL for this sample. The PSI was 0.82, and this showed the scale to have adequate levels of internal reliability. Investigation of the individual item fit showed that item 2 ($P<0.001$) and item 7 ($P=0.037$) misfit the Rasch model. Items 6–9 had disordered response thresholds. Items 2 ($P=0.003$), 4 ($P=0.033$), and 6 ($P=0.047$) showed significant uniform DIF by gender, and items 2 ($P=0.009$), 4 ($P=0.050$), and 8 ($P=0.020$) exhibited significant nonuniform DIF by gender. Items 5 ($P=0.040$) and 10 ($P=0.002$) showed uniform DIF by age. Items were again clustered at the center of the HRQL spectrum with too few mild items.

Atopic dermatitis

The overall fit statistics showed that the DLQI fit the Rasch model for the AD patients (item–trait interaction = 0.460). Despite this, several problems with the measurement properties of the scale were identified. The PSI was low, indicating that the scale did not have adequate internal reliability. In addition, the item residual statistics were indicative of misfit to the Rasch model, suggesting problems with the items. A single item (item 4, $P=0.048$) showed misfit to the model. Items 2 and 4–8 had disordered response thresholds. The

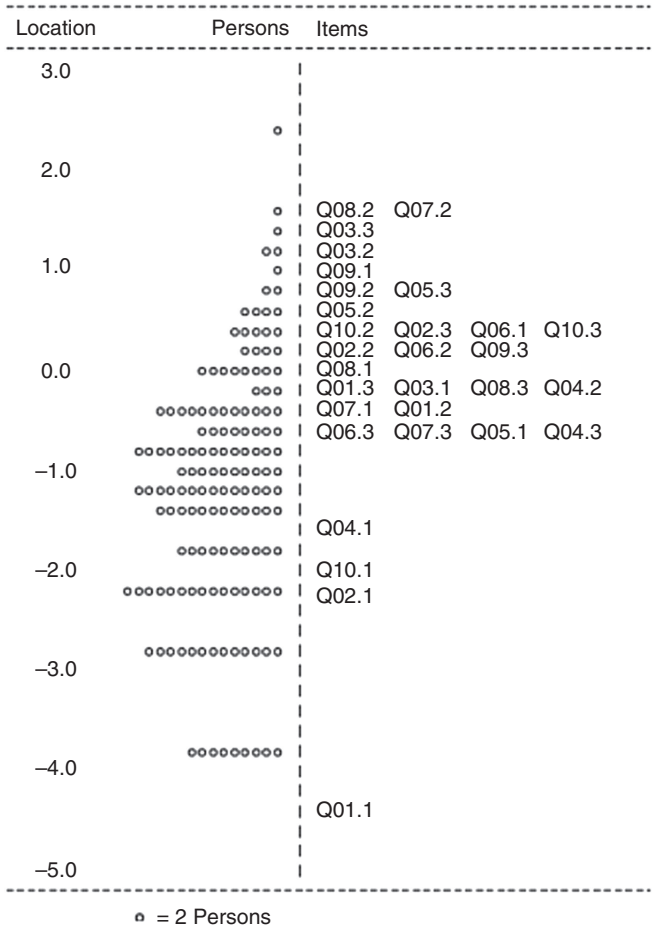


Figure 3. Logit scale. In Rasch analysis, the level of difficulty of items and the ability of persons are measured using the logit scale. The figure shows the logit position of each item response and the logit position of each person.

response thresholds were also poorly defined in the other items, suggesting that a complete restructuring of the response options is necessary. Items 4 ($P=0.010$) and 7 ($P=0.043$) showed uniform DIF by gender, and item 6 ($P=0.012$) exhibited nonuniform DIF by gender. Items 2 ($P=0.010$), 4 ($P=0.020$), 7 ($P<0.001$), and 10 ($P=0.028$) showed uniform DIF by age, and item 7 (<0.001) showed nonuniform DIF by age. Items were again clustered at the center of the HRQL spectrum with too few mild items.

DISCUSSION

A growing number of studies are using Rasch analysis to assess the measurement properties of widely used HRQL measures, such as the SF-36 and the Hospital Anxiety and Depression Scale (Raczek *et al.*, 1998; Smith *et al.*, 2006; Dallmeijer *et al.*, 2007; Pallant and Tennant, 2007). Review of these studies shows that there are often significant problems with the scales, and that several changes would be necessary to improve their measurement properties. Where significant problems are found, this indicates that the scales are not valid in their standard format and may only produce valid scores after the raw scores have been transformed using Rasch analysis. Researchers should consider carefully before making substantial changes to existing measures, as this may violate the theoretical structure of the scale and detract from the developers' intentions that the scales should be simple to use and score.

Fit to the Rasch model for the combined sample

To our knowledge, no previous study has assessed the measurement properties of the DLQI in both psoriasis and AD populations. The study found that the DLQI failed to meet fundamental measurement requirements, indicating that it does not measure a single unidimensional construct. As a consequence, total scores on the measure do not provide valid information about the respondent.

Although the internal reliability (as indicated by the PSI) of the DLQI was good, several problems with the measurement properties of the scale were identified. Three items misfit the Rasch model, indicating that respondents were not answering these items in the way the model would predict. DIF analysis showed that several items showed DIF associated with age, gender, and/or disease. The evidence of DIF by disease indicates that DLQI scores for patients with psoriasis and AD should not be compared. In addition, scores for these different patient groups should not be combined during clinical trials or research studies.

Too few items covered the mild end of the measurement scale, indicating that the DLQI will not be responsive to change in individuals with mild disease. Response thresholds were disordered for five items and poorly defined for the remainder. The greatest source of disorder resulted from the response options ''a lot'' and ''very much''. Patients clearly had difficulty distinguishing between these categories that are conceptually similar. Item frequencies also showed that large proportions of the samples answered ''not relevant'' to several items. ''Not relevant'' responses are given the same score as the ''not at all'' response, according to the DLQI scoring guidelines (Finlay and Khan, 1994). This scoring method presents a problem as individuals who responded ''not relevant'' may actually have had severe illness. Future research using Rasch analysis is needed to investigate the impact of this anomaly.

Overall fit to the Rasch model during refinement

Several changes were made to the DLQI in an attempt to improve the measurement properties of the scale. An acceptable level of fit to the model was eventually achieved in Analysis 7 after substantial changes had been made. These included altering the response options for 5 of the 10 items, and splitting 3 of the items in order to reduce the levels of DIF. Although this produced the best compromise, some problems with the measurement properties of the scale remained. The item fit residuals still indicated that there were problems with the items, which appeared to be caused by the misfit of item 2 and low-level DIF by age and/or gender and/or disease in several items. Furthermore, there were still too few items measuring mild illness. As several major changes to the scale were required to improve its measurement properties, it is clear that the DLQI does not provide valid

measurement. It is also likely that additional DIF would arise from using the measure with patients who have different types of skin disease. Analysis of responses to the measure would also become unwieldy, negating the developers' intention that the scale should be short and easy to apply.

Fit to the Rasch model for the separate psoriasis and AD samples

The DLQI misfit the Rasch model for the psoriasis sample, indicating that it does not measure a unidimensional construct for psoriasis patients. The DLQI did fit the Rasch model for the AD sample, but showed several measurement problems. Similar problems were identified for the two groups separately as with the combined sample. This indicated that both disease groups contributed to the misfit to the Rasch model of the overall sample. In both diseases, items misfit the Rasch model and exhibited DIF, response thresholds were disordered, and there were too few mild items. Although the DLQI did fit the Rasch model for the AD group, the measurement properties of the scale were still poor. For example, the internal reliability (PSI) for the measure was poor, and the item residual statistics indicated that there were problems with the items. In addition, 6 of the 10 items for the AD sample had disordered response thresholds, and a complete restructuring of the response format seemed appropriate.

The results of this study support two previous studies that investigated the scaling properties of the DLQI using Rasch analysis (Nijsten *et al.*, 2006a, 2007). The first study found that the DLQI misfit the Rasch model for both psoriasis and AD samples. Both studies identified problems with individual items in terms of misfit, DIF by age and gender, problems with the response format, and with the measurement range of the questionnaire. The second study also investigated the cross-cultural equivalence of the DLQI using DIF analysis. This study involved 450 patients with psoriasis from five European countries and the United States. Results showed that the DLQI misfit the Rasch model for the combined-country data and for three of the countries separately. In addition, each of the 10 items in the DLQI showed DIF by country, indicating that the DLQI was working differently in each country and that, consequently, international clinical trial data should not be combined.

A previous study, using confirmatory factor analyses, reported that the items in the DLQI form a higher-order unidimensional factor for psoriasis patients (Mazzotti *et al.*, 2005). However, insufficient detail regarding model fit statistics for the individual items was provided, making meaningful comparisons with the present study difficult. The detailed analyses of the present study have shown several problems with the scale beyond the simple question of overall fit to the model. The advantages of the Rasch model over other approaches (such as factor analysis and two-parameter Item Response Theory models) have been discussed in detail (Wright, 1996; Prieto *et al.*, 2003; Andrich, 2004; Tennant *et al.*, 2004; Waugh and Chapman, 2005).

It is likely that the problems identified with the DLQI are due to in-built scale development weaknesses. During the development of the DLQI, an overarching aim was to develop a scale that would fit on one side of paper (Finlay and Khan, 1994). Although this means that the DLQI is short and easy to administer, this resulted in a scale that was unable to measure the full range of HRQL impact associated with dermatological disease. Many of the items are also ambiguous in nature. Several include two or more ideas in the same question, forming ''double-barreled'' questions. For example, item 3 asks how the patient's skin has interfered with ''shopping'', looking after the ''home'', or ''garden''. With such an item, patients may respond to different parts of the question. This causes problems for the measure as the different parts of the question represent different levels of difficulty. Researchers have noted the relation between double-barreled questions and DIF (Hambleton, 2006). In addition, the phraseology of some of the items, e.g., ''leisure activities'', is likely to present problems for some respondents.

Alternative measures for health outcome assessment in dermatological conditions are available, which have been developed with the benefit of Rasch analysis. A subset of the items in the Skindex (the Skindex-17) form a generic dermatological measure that includes unidimensional measures of psychosocial functioning and symptoms (Chren *et al.*, 1997; Nijsten *et al.*, 2006b). Two other disease-specific QoL measures are also available. The Psoriasis QoL scale (McKenna *et al.*, 2003) and the Quality of Life Index for Atopic Dermatitis (Whalley *et al.*, 2004) are unidimensional measures of QoL that used the needs-based QoL model (Hunt and McKenna, 1992) and were developed using Rasch analysis. These latter two measures are disease specific and assess a different outcome from the DLQI (QoL rather than HRQL).

Strengths and limitations of the study

This is one of the first studies to report on the Rasch analysis of the DLQI, and, to our knowledge, the investigation of the equivalence of the DLQI for different dermatological conditions is previously unreported. Study data were derived from two separate samples: psoriasis and AD. The psoriasis sample was recruited from hospital records, whereas the AD sample was identified through a self-help organization. The different sampling methods are dealt with within the Rasch framework, which is sample independent, as all patients are measured on the same underlying construct. For the analyses of DIF by dermatological condition, patients with the same level of the underlying construct (HRQL impairment) are compared, rather than comparing the two complete samples.

The total sample of 292 provided an excellent number for Rasch parameter estimates as it was large enough to give 99% confidence that the parameter estimates were within half a logit of the stable value (Linacre, 1994). The two separate samples ($n = 141$) provided 95% confidence that the estimates were within half a logit of the stable value.

This study illustrated the use of the DLQI with psoriasis, AD and combined psoriasis, and AD samples only. The DLQI is also used with several other dermatological conditions and further research could determine whether the scale works adequately with such conditions.

Conclusions

The DLQI has been described as a first-generation HRQL measure in dermatology as it was developed without the benefit of modern Item Response Theory techniques (Nijsten *et al.*, 2007). The fact that the DLQI has several measurement problems is therefore not surprising. It is recommended that a new measure of functional limitations in dermatology be developed, which benefits from modern scaling techniques. The DLQI should not be used to compare scores of psoriasis and AD patients on the basis of the findings of the current study. It may also not be valid to compare scores for groups of patients whose profile differs in terms of age or gender. The use of the DLQI for deciding which patients should receive biological treatments in the United Kingdom is of concern, given the lack of support for the measurement properties of the scale.

MATERIALS AND METHODS

Design and data

Rasch analyses were based on two available data sets. AD and psoriasis patients in the United Kingdom completed the DLQI in postal surveys used to validate new psoriasis (Psoriasis QoL; McKenna *et al.*, 2003) and AD (Quality of Life Index for Atopic Dermatitis; Whalley *et al.*, 2004) quality of life questionnaires. The psoriasis sample was recruited from the Manchester Psoriasis Service database at Salford Royal Hospital, Greater Manchester, UK. The AD sample was recruited through the National Eczema Society. The original AD sample included 202 females (70.6%) and 84 males (29.4), but a random sample (generated via SPSS 16 random sample selection generator) of 73 males and 73 females was selected to match the number of males and females in the psoriasis sample. This was because uneven sample sizes can distort analysis of variance analyses used for the investigation of DIF (Wright and Tennant, 1996; Hambleton, 2006). Information was collected describing patient's duration of illness, marital status, self-perceived general health (very good/good/fair/poor), and self-perceived disease severity (mild/moderate/quite severe/very severe).

Rasch analyses were conducted on the overall group and then on the psoriasis and AD sample separately.

Dermatology Life Quality Index

The DLQI is a self-report 10-item questionnaire assessing the impact of skin disease on the patient over the previous week. The items cover symptoms, treatment, activity limitations, and emotional reactions to having a skin disease. Nine items have four response options: ''Not at all'', ''A little'', ''A lot'', and ''Very much'', whereas item 7 first asks whether work or study has been prevented and then (if ''No'') to what degree the skin condition has been a problem at work/study (''A lot'', ''A little'', or ''Not at all''). Eight of the items also have a ''Not relevant'' option that is scored ''0'', indicating no problem. Individual item scores are summed to derive a total DLQI score that can range from 0 to 30, with higher scores indicating worse HRQL.

Analysis

Demographic information. Tests were conducted to compare the AD and psoriasis samples in terms of their length of illness (Mann–Whitney *U*-test), DLQI scores (Man–Whitney *U*-test), general health (χ^2 analysis), and perceived illness severity (χ^2). Non-parametric statistical analyses were conducted throughout because of the non-normal distribution/ordinal nature of the data.

Item frequency analysis. Item response frequencies were investigated before the data were Rasch analyzed.

Rasch analysis. The Rasch model is a simple logistic latent trait Item Response Theory model. Rasch analysis places response data for each individual and each item on the same spectrum of severity (logit scale). According to the model, the probability that an individual will respond in a certain way to a particular item is a logistic function of the relative distance between the item location (parameter) and the person location (parameter), and only a function of these two factors. Persons and items are plotted on the same logit scale on the basis of the difference in their location on the underlying spectrum. This difference governs the probability of the expected response for a person, of a given severity, on a question of a given severity. If the observed data do not deviate significantly from the expected responses, then the items fit the Rasch model.

Rasch analysis also provides the opportunity to assess the measurement properties of a questionnaire in a number of additional ways:

- Where questionnaire items have more than two response options, Rasch analysis allows the functioning of the response format to be examined.
- As Rasch analysis also places each individual and each item on the same logit spectrum, information is provided on the order of each item in terms of its severity (and therefore the relative impact of the item in terms of, say, HRQL impact) and whether the items as a whole are well targeted to respondents.
- DIF represents instability in the order of severity of items and indicates that the scale may not work in the same way in subgroups of individuals (for example, diagnosis subgroups, age, or gender) who share the same level of trait being measured (Holland and Wainer, 1993). Uniform DIF occurs when one group is more or less likely to affirm an item at each level of severity. Nonuniform DIF occurs when there is inconsistency in which group is more likely to affirm an item for different levels of severity.

The Rasch Unidimensional Measurement Model 2020 (Andrich *et al.*, 2005) program was used for the analyses.

Analyses conducted were consistent with published guidelines (Tennant and Gonaghan, 2007). Before analyses were conducted, it was necessary to determine whether the rating scale (Andrich, 1978) or partial credit model (Masters, 1982) was most appropriate. Both of these approaches use the Rasch model, but differ slightly in their mathematics. The rating scale model is more stringent in its requirements as the distance between response thresholds has to be uniform across all items. The likelihood ratio test is used to identify the most appropriate model.

Internal reliability was analyzed using the PSI. The PSI is indicative of the power of the construct to distinguish among respondents. A PSI score of 0.70 is the minimum acceptable level. The overall scale fit to the model was examined by reference to the

overall item–trait interaction χ^2-fit value. A significant χ^2-statistic ($P<0.05$) indicates that there is a real deviation of the scale from the expected pattern and a lack of fit to the Rasch model. Overall fit of the data is also investigated via Item and Person interaction statistics. These assessments measure the extent to which observed item and person estimates deviate from the expected. The mean location of the items is always anchored at 0. Within this function, both Person fit residual and Item fit residual statistics are transformed by Rasch Unidimensional Measurement Model to approximate a *Z*-score; this represents a standardized normal distribution. Therefore, when the data fit the model, the overall distribution statistics for Item fit and Person fit should have a mean of ~ 0 and a standard deviation of ~ 1.

Individual item fit statistics were also investigated via χ^2-fit statistics. A significant χ^2-fit statistic ($P<0.05$) indicated misfit to the model. Individual item fit residuals were also consulted. These should fall within ± 2.5 if all individuals responded in the anticipated way. High negative residuals were indicative of overfit to the scale (item redundancy), and high positive residuals were indicative of misfit to the scale (low association with the scale score).

The functioning of the response options was investigated via response thresholds. The thresholds represent the borders between adjacent ordinal response options and should increase in a logical way. If response options are disordered, the response categories are not working in the intended way.

An analysis of variance of standardized residuals was carried out to examine DIF by gender (males vs females), age group (below the median age vs median age and above), and disease (AD versus psoriasis). If the analysis of variance *P*-value was <0.05, then that item was considered to exhibit DIF.

Targeting of the items to respondents was assessed by examining the item map. These show the ordering of both persons and items on the same logit scale and indicate whether the items in the scale are well matched to the respondents.

If a scale is unidimensional, it should be free from item dependency: similarity in items such that answers to one item have too strong an influence over answers to other items (Tennant and Gonaghan, 2007). This is assessed by inspecting the residual correlation matrix for pairs of items with correlations exceeding 0.3.

Refinement of the measure

Rasch analysis can be used in an attempt to correct for problems with the measurement properties of a scale. The following approaches can be taken:

- Individuals that misfit the Rasch model (fit residuals $> \pm 2.5$) can be removed, as these cases do not respond in a way predicted by the model.
- The response options can be altered by collapsing thresholds between adjacent response options if they are working incorrectly.
- Items can be deleted if they misfit the Rasch model.
- Items exhibiting uniform DIF can be split to allow them to have different parameters for each group (for example, males and females in the case of DIF by gender). Essentially, this method allows an item to become two separate items with different item locations.

The present study attempted to improve the measurement properties of the DLQI by applying these methods.

CONFLICT OF INTEREST

Galen Research is the developer of several dermatology patient-reported outcomes. However, these assess true quality of life rather than HRQL as assessed by the DLQI. In this respect, they are complementary rather than competing outcome measures.

ACKNOWLEDGMENTS

We thank the patients who completed the DLQI, Professor CEM Griffiths from Salford Royal Hospital, and the UK National Eczema Society.

REFERENCES

Andrich D (1978) Rating formulation for ordered response categories. *Psychometrika* 43:561–73

Andrich D (2004) Controversy and the Rasch model: a characteristic of incompatible paradigms? *Medical Care* 42(Suppl S1):1–7

Andrich D, Sheridan B, Ludo G (2005) *Rasch Unidimensional Measurement Models: A Windows-Based Item Analysis Program Employing Rasch Models (RUMM2020).* Perth, Australia: Murdoch University

Basra MKA, French R, Gatt RM *et al.* (2008) The dermatology Life Quality Index 1994–2007: a comprehensive review of validation and clinical results. *Br J Dermatol* 159:997–1035

Brodersen J, Thorsen H, Kreiner S (2010) Consequences of screening in lung cancer: development and dimensionality of a questionnaire. *Value Health* 13:601–12

Chren MM, Lasek RJ, Flocke SA *et al.* (1997) Improved discriminative and evaluative capability of a refined version of Skindex, a quality of-life instrument for patients with skin diseases. *Arch Dermatol* 133: 1433–40

Dallmeijer AJ, dr Groot V, Roorda LD *et al.*, FuPro Study Group (2007) Cross-diagnostic validity of the SF-36 physical functioning scale in patients with stroke, multiple sclerosis and amyotrophic lateral sclerosis: a study using Rasch analysis. *J Rahabil Med* 39:163–9

Finlay AY, Khan GK (1994) Dermatology Life Quality Index (DLQI)-a simple practical measure for routine clinical use. *Clin Exp Dermatol* 19: 210–6

Hambleton RK (2006) Good practices for identifying differential item functioning. *Med Care* 44:S182–8

Holland PW, Wainer H (eds) (1993) *Differential Item Functioning.* New Jersey: Lawrence Erlbaum Associates

Hunt SM, McKenna SP (1992) The QLDS: a scale for the measurement of quality of life in depression. *Health Policy* 22:307–19

Lewis V, Finlay AY (2004) Ten years experience of the Dermatology Life Quality Index (DLQI). *J Investig Dermatol Symp Proc* 9:169–80

Linacre JM (1994) Sample size and item calibration stability. *Rasch Measurement Trans* 7:4. p.328

Ludwig MW, Oliveira Mda S, Muller MC *et al.* (2009) Quality of life and site of the lesion in dermatological patients. *An Bras Dermatol* 84: 143–50

Luquet C, Chau N, Guillemin F *et al.* (2001) A method for shortening instruments using the Rasch model. Validation on a hand functional measure. *Rev Epidemiol Sante Publique* 49:273–86

Masters G (1982) A Rasch model for partial credit scoring. *Psychometrika* 47:149–74

Mazzotti E, Barbaranelli C, Picardi A *et al.* (2005) Psychometric properties of the Dermatology Life Quality Index (DLQI) in 900 Italian patients with psoriasis. *Acta Derm Venereol* 85:409–13

McKenna SP, Cook SA, Whalley D *et al.* (2003) Development of the PSORIQoL, a psoriasis-specific measure of quality of life designed for use in clinical practice and trials. *Br J Dermatol* 149:323–31

National Institute for Health and Clinical Excellence (2008a) Infliximab for the treatment of adults with psoriasis. *NICE Technology Appraisal Guidance 134*

National Institute for Health and Clinical Excellence (2008b) Adalimumab for the treatment of adults with psoriasis. *NICE Technology Appraisal Guidance 146*

National Institute for Health and Clinical Excellence (2009) Ustekinumab for the treatment of adults with moderate to severe psoriasis. *NICE Technology Appraisal Guidance 180*

Nijsten T, Meads DM, de Korte J *et al.* (2007) Cross-cultural inequivalence of dermatology-specific health related quality of life instruments in psoriasis patients. *J Invest Dermatol* 127:2315–22

Nijsten T, Meads DM, McKenna SP (2006a) Dimensionality of the dermatology life quality index (DLQI): a commentary. *Acta Derm Venereol* 86:284–5; author reply 285–6

Nijsten TE, Sampogna F, Chren MM *et al.* (2006b) Testing and reducing skindex-29 using Rasch analysis: Skindex-17. *J Invest Dermatol* 126:1244–50

Pallant JF, Tennant A (2007) An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *Br J Clin Psychol* 46:1–18

Papoutsaki M, Chimenti MS, Costanzo A *et al.* (2007) Adalimumab for severe psoriasis and psoriatic arthritis: an open-label study in 30 patients previously treated with other biologics. *J Am Acad Dermatol* 57:269–75

Potocka A, Turczyn-Jabtońska K, Kieć-Swierczyńska M (2008) Self-image and quality of life of dermatology patients. *Int J Occup Med Environ Health* 21:309–17

Potocka A, Turczyn-Jabloñska K, Merecz D (2009) Psychological correlates of quality of life in dermatology patients: the role of mental health and self-acceptance. *Acta Dermatovenerol Alp Panonica Adriat* 18:53–8

Prieto L, Alonso J, Lamarca R (2003) Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual Life Outcomes* 1:27

Quandt SA, Schulz MR, Vallejos QM *et al.* (2008) The association of dermatologist-diagnosed and self-reported skin diseases with skin-related quality of life in Latino migrant farmworkers. *Int J Dermatol* 47:236–41

Raczek AE, Ware JE, Bjorner JB *et al.* (1998) Comparison of Rasch and summated rating scales constructed from the SF-36 physical functioning items in seven countries: results from the IQOLA project. *J Clin Epidemiol* 51:1203–14

Rasch G (1960) (Reprinted 1980) *Probabilistic Models for Some Intelligence and Attainment Tests*. Chicago: University of Chicago Press

Schmitt J, Heese E, Wozel G *et al.* (2007) Effectiveness of inpatient treatment on quality of life and clinical disease severity in atopic dermatitis and psoriasis vulgaris—a prospective study. *Dermatology* 214:68–76

Smith AB, Wright EP, Rush R *et al.* (2006) Rasch analysis of the dimensional structure of the Hospital Anxiety and Depression Scale. *Psychoonconology* 15:817–27

Smith CH, Anstey AV, Barker JN *et al.* (2005) British Association of Dermatologists Guidelines for use of biological interventions in psoriasis 2005. *Br J Dermatol* 153:486–97

Smith CH, Anstey AV, Barker JN *et al.* (2009) British Association of Dermatologists' guidelines for biologic interventions for psoriasis 2009. *Br J Dermatol* 161:987–1019

Tennant A, Gonaghan PG (2007) The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one should one look for in a Rasch paper? *Arthritis Rheum* 57:1358–62

Tennant A, McKenna SP, Hagell P (2004) Application of Rasch analysis in the development and application of quality of life instruments. *Value Health* 7(Suppl 1):S22–6

Waugh RF, Chapman ES (2005) An analysis of dimensionality using factor analysis (true-score theory) and Rasch measurement: what is the difference? Which method is better? *J Appl Meas* 6:80–99

Whalley D, McKenna SP, Dewar AL *et al.* (2004) A new instrument for assessing quality of life in atopic dermatitis: international development of the Quality of Life Index for Atopic Dermatitis (QoLIAD). *Br J Dermatol* 150:274–83

Wright BD (1996) Comparing Rasch measurement and factor analysis. *Structural Equation Modelling* 3:3–24

Wright BD, Tennant A (1996) Sample size again. *Rasch Measurement Transactions* 9:4. (p468) http://www.rasch.org/rmt/rmt74m.htm accessed 30 November 2009